Identifying the inheritable component of human thymic T cell repertoire generation in monozygous twins

Heikkilä Nelli[1,2], Vanhanen Reetta[1,2], Yohannes Dawit A.[1,3], Saavalainen Päivi[1,3], Meri Seppo[1,2], Jokiranta T. Sakari[2], Jarva Hanna[2], Mattila Ilkka P.[4], Hamm David[5], Sormunen Silja[6], Saramäki Jari[6], Arstila T. Petteri[1,2]

1) Research Programs Unit, Translational Immunology, University of Helsinki, Finland
2) Medicum, Department of Bacteriology and Immunology, and Research Programs Unit, Immunobiology, University of Helsinki, Finland
3) Medicum, Department of Medical and Clinical Genetics, University of Helsinki, Finland
4) Department of Pediatric Cardiac and Transplantation Surgery, Hospital for Children and Adolescents, Helsinki University Central Hospital, Finland
5) Adaptive Biotechnologies, Seattle, Washington
6) Department of Computer Science, Aalto University, Finland

Keywords: TCR repertoire, twins, genetics, thymic selection, CDR3

Corresponding author:

Nelli Heikkilä

Research Programs Unit, Translational Immunology, PB21, 00014 University of Helsinki, Finland

e-mail: nelli.heikkila@helsinki.fi

telephone: +358 29 41 26373

fax: +358 29 41 26382

Abbreviations:

D: diversity gene segment

J: joining gene segment

JI: Jaccard index

TCR: T cell receptor

V: variable gene segment

Text

A basic determinant of adaptive immune responses is T cell antigen receptor (TCR) repertoire, generated in a stochastic fashion during thymic T cell development(1). Heterodimeric TCRs consist of α and β chains assembled in somatic recombination of gene segments and are subjected to rigorous selection steps that eliminate nonfunctional and potentially self-reactive receptors(2). Despite the stochastic nature of TCR repertoire generation, inheritable factors may influence both the recombination and selection, as genetic factors determine the gene segment composition, the enzymatic machinery, and to some extent also the TCR selecting ligandome. Inheritability of circulating TCR repertoires has been assessed in monozygous twins, identifying a distinct genetic bias in TCR gene segment usage but very little genetic influence on junctional sequences and sharing of clonotypes(3–5). Higher junctional sharing between twins has been shown among high frequency clonotypes in the periphery and the authors attributed this to the genetic impact in the generation of immunological memory(3). Another analysis, based on a model comparing sharing in computationally predicted and actually sequenced repertoires, showed higher sharing in circulating sequences than predicted by the model between twins but not between unrelated individuals. Because part of the shared sequences showed low generation probabilities, the authors attributed the excess sharing to the persistence of fetal clones generated by one twin and physically shared via the common placental circulation during gestation(5).

Cardiac surgery performed on a pair of monozygous twins offered us a rare opportunity to address the role of genetic factors in generating TCR repertoire in the human thymus. We studied pediatric thymus samples from the twins (Twins A and B; 8 months) and three unrelated controls (Controls A, B and D; 4-8 months), all immunologically healthy. Flow cytometric analysis showed normal distribution of CD3, CD4 and CD8 markers in thymus samples from Twin B and all controls (SI, Fig. 1). Because we recognized the siblings as monozygous twins only at the time of Twin B's surgery, the sample from Twin A was limited to a single frozen pellet of 10 million cells, making flow cytometry unfeasible. Despite the

obvious limitations in sample collection, and the sample size of only one twin pair, the setting provides a unique opportunity to study TCR generation at its very primordial level in postnatal thymus.

Ten million thymocytes from each sample were sequenced independently for TCRα and TCRβ chains. Two independent aliquots from Twin B were sequenced. An average of 13 and 58 million total reads were obtained for β and α TCRs, respectively, corresponding to 1.3 and 6.5 million unique nucleotide sequences. The fraction of inframe sequences was 79% for β and 31% for α repertoire. Productivity of α sequences varied minimally between subjects, whereas the β chain productivity was heterogeneous between individuals and more similar in twins than in unrelated individuals (SI Table 1). Rubelt et al. have reported that genetics affect the CDR3 length and number of non-templated insertions in the circulating repertoire, but no such effect was seen in our data(4).

A genetic bias in V and J gene usage, previously reported in circulating T cells(6), was also evident in the thymus, for both β and α sequences. Principal component analysis of gene segment usage for V genes, J genes and VJ combinations showed clear clustering of the twins in the inframe repertoire compared along the first two principal components (SI Fig. 2). Similar results were obtained when analyzing the nonproductive (out-of-frame) repertoire. Since these sequences remain unexpressed, the nonproductive repertoire avoids direct shaping by thymic selection and allows the analysis of preselected repertoire. In contrast, Zvyagin et al. reported no genetic influence in nonproductive Jα usage in periphery and suggested that this reflected the effects of thymic selection(3). We note that we used genomic DNA as starting material for analysis, while Zvyagin et al. used mRNA. The TCR mRNA content per cell is known to vary widely and in particular the decay of nonproductive transcripts may be rapid(7). This may affect the clonal representation of some sequences within the library.

To assess the genetic input in the generation of junctional sequences, we measured the fraction of shared clonotypes in the repertoire by calculating Jaccard index (JI), i.e. the intersection of two sequence sets

divided by their union. In the nonproductive β sequences, reflecting the preselection repertoire, the twins showed a clearly higher overlap than unrelated individuals (Fig. 1A). However, in the TCRβ inframe nucleotide and amino acid repertoires the sharing between twins was only marginally higher than between unrelated individuals. The repeat samples from Twin B scored the highest overlap rates independent of sequence productivity, indicating sufficient sequencing depth. In contrast, for α repertoire the sharing in the repeat samples from Twin B was similar to the sharing between Twin A and Twin B. This may be due to the higher diversity of TCRα than TCRβ repertoire in the thymus(8). Interindividual comparison of the TCRα repertoire showed a small but consistently higher sharing for twin samples compared with unrelated samples in both nonproductive and inframe nucleotide sequences and amino acid chains (Fig. 1A).

Because the observed overlap in TCR repertoires is susceptible to the sample size, we recalculated JIs after random resampling of each sample to the size of the smallest dataset in both TCRβ and TCRα repertoires. The difference between twins and unrelated individuals was still evident for nonproductive TCRβ sequences (Fig. 1B) but a small difference was now also detected for inframe nucleotide sequences. However, in TCRβ amino acid repertoire the JI distributions were overlapping. In the TCRα repertoire the resampled twin samples clustered together both in nonproductive and in inframe nucleotide sequences. Also, in the TCRα amino acid repertoire the twins had a slightly higher JI than any of the unrelated samples, although the difference was small (Fig. 1B).

To further test genetic input in the generation of junctional sequence motifs, we analyzed the frequencies of all possible tetranucleotide combinations in the CDR3 region. The interindividual similarity of these frequency distributions was measured with cosine distance. For TCRβ the similarity of CDR3 tetranucleotide distributions was higher in twins than unrelated individuals (Fig. 2A), and for TCRα the difference was even more distinct. However, since both ends of the CDR3 are often germ-line-encoded, these results might have been skewed by the higher sharing of gene elements found in the twins. To

exclude this possibility, we randomly selected six VJ combinations for tetranucleotide frequency analysis, obtaining essentially similar results (Fig 2B).

In addition to the overlap of unique clonotypes, the similarity of two repertoires is defined by frequencies of different clonotypes. To test whether the same clones are enriched in different individuals, we correlated the shared clonotype sizes in two individuals. TCRβ repertoire showed no clear correlation in the clonotype sizes whereas in TCRα the correlation was obvious and much more marked between twins than unrelated individuals (Fig 2C).

In summary, our analysis of thymic TCR repertoire confirm the previously observed genetic bias in VJ usage but additionally show a detectable genetic element in the generation of junctional CDR3 sequences. The genetic signal seems to be imprinted in the somatic recombination and diluted by intrathymic selection, suggesting that stochastic factors dominate the selection. Moreover, the genetic signal seems to be retained better in the expressed α than in the β repertoire, as shown by both sequence and tetranucleotide motif analyses.

Acknowledgements: We thank Tamás Bazsinka from University of Helsinki for expert technical assistance, and the families who participated in the study. The study has been funded by the University of Helsinki Research Foundation, Emil Aaltonen Foundation, Biomedicum Helsinki Foundation and MD PhD program of the University of Helsinki

Conflict of interest: The authors declare no commercial or financial conflicts of interest except for David Hamm who is employed by Adaptive Biotechnologies, a company providing commercial TCR sequencing services and products.

References

1. **Jaeger S. et al.** Immunology. 2013. **139**: (2)141–50.
2. **Klein L. et al.** Nat. Rev. Immunol. 2014. **14**: (6)377–91.
3. **Zvyagin I.V. et al.** Proc. Natl. Acad. Sci. U. S. A. 2014. **111**: (16)5980–5.
4. **Rubelt F. et al.** Nat. Commun. 2016. **7**: 11112.
5. **Pogorelyy M.V. et al.** PLoS Comput. Biol. 2017. **13**: (7)e1005572.
6. **Quirós Roldan E. et al.** Immunogenetics. 1995. **41**: (2-3)91–100.
7. **Chang Y.-F. et al.** Annu. Rev. Biochem. 2007. **76**: 51–74.
8. **Vanhanen R. et al.** Mol. Immunol. 2016. **76**: 116–22.

Fig 1. Sequence overlap measured with Jaccard index between twins and unrelated samples in total repertoires(A) and in repertoires resampled to be equal in size(B). Comparison of twin samples is shown as circles, the open circle indicating the repeat samples from Twin B. Comparison of unrelated samples is shown as squares and the mean value indicated with horizontal bar. Note the discontinuous scale in graphs for TCRβ repertoire.

Figure 2. The similarity of tetranucleotide motif frequency distributions is shown as a heatmap indicating cosine distances in total repertoire(A) and with six randomly selected VJ combinations(B). The scale is indicated separately for each plot, value zero indicating exactly similar distributions. The correlation between clonotype sizes in shared repertoires(C) is shown between twins, and between two representative unrelated individuals (Twin A–Control A, Control A–Control B). The correlations for TCRβ and TCRα are shown in upper and lower panels, respectively.